

Groups, Information Theory and Einstein's Likelihood Principle

Gabriele Sicuro*

Centro Brasileiro de Pesquisas Físicas, Rua Dr. Xavier Sigaud, 150, 22290-180, Rio de Janeiro, Brazil

Piergiulio Tempesta†

*Departamento de Física Teórica II (Métodos Matemáticos de la física),
Facultad de Físicas, Universidad Complutense de Madrid, 28040 – Madrid,
Spain, and Instituto de Ciencias Matemáticas (CSIC-UAM-UC3M-UCM),
C/ Nicolás Cabrera, No 13–15, 28049 Madrid, Spain*

(Dated: April 7, 2016)

We propose a unifying picture where the notion of generalized entropy is related to information theory by means of a group-theoretical approach. The group structure comes from the requirement that an entropy be well defined with respect to the composition of independent systems, in the context of a recently proposed generalization of the Shannon–Khinchin axioms. We associate to each member of a large class of entropies a generalized information measure, satisfying the additivity property on a set of independent systems as a consequence of the underlying group law. At the same time, we also show that Einstein's likelihood function naturally emerges as a byproduct of our informational interpretation of (generally nonadditive) entropies. These results confirm the adequacy of composable entropies both in physical and social science contexts.

The study of the relations among statistical mechanics, information theory and the notion of entropy is at the heart of the science of complexity, and in the last decades has been widely explored. After the seminal works of Shannon [1] and Khinchin [2] on the foundations of Information Theory, Jaynes [3] re-formulated Boltzmann–Gibbs (BG) statistical mechanics as a statistical inference theory, where all fundamental equations are consequences of the maximum entropy principle applied to the *BG entropy* S^{BG} . Subsequently, Rényi [4, 5] introduced a generalized measure of information, now called the *Rényi entropy* S_α^{R} , that depends on a real parameter α having the BG entropy as a particular case in the $\alpha \rightarrow 1$ limit. The S_q^{T} entropy, introduced by Havrda and Charvát [6] and Tsallis [7], has been the prototype of the nonadditive entropies studied in the last decades [8–13]. These functionals are generalizations of the BG entropy; they depend on one or more parameters, in such a way that the BG entropy is recovered as a particular limit. Generalized entropies have been successfully adopted for the study of both classical and quantum systems. Rényi's entropy, for example, plays a central role in information theory and in the study of multifractality [14]; along with the von Neumann entropy, it has been also used extensively in the evaluation of the entanglement entropy of quantum systems [15–18].

The study of new entropic forms has led to a new flow of ideas regarding the old problem of the probabilistic versus the dynamical foundations of the notion of entropy. It is well known [19] that Einstein's approach was very different with respect to the probabilistic methodology of Boltzmann (which eventually emerged as the predominant one). Indeed, Einstein argued that the probabilities of occupation of the various regions of the phase space associated with a physical systems cannot be pos-

tulated *a priori*. Instead, only a knowledge of dynamics, obtained by solving the equations of motion, could provide this information. For this reason, in his theory of fluctuations, Einstein [20] introduced the likelihood function $\mathcal{W} \propto \exp(S^{\text{BG}})$ as a fundamental statistical quantity (for the sake of simplicity, here and in the following, we put $k_{\text{B}} \equiv 1$, k_{B} being the Boltzmann constant). He observed that, by composing two independent systems \mathcal{A} and \mathcal{B} , the fundamental relation

$$\mathcal{W}(\mathcal{A} \cup \mathcal{B}) = \mathcal{W}(\mathcal{A})\mathcal{W}(\mathcal{B}) \quad (1)$$

holds. Equation (1) is epistemologically crucial: it expresses the fact that the physical description of system \mathcal{A} does not depend on the physical description of system \mathcal{B} , and *vice versa*. Moreover, it is related to the additivity requirement of the information content of independent systems, as will be explained below.

The likelihood function has a clear physical meaning. It provides the number of accessible configurations in the entire space of possible configurations. In many circumstances, this number is exponential in the size N of the system; therefore we can write $\mathcal{W} \propto \exp(N\Sigma)$, where Σ is an (adimensional) entropy density. Mézard and Parisi [21] introduced this quantity in the study of disordered systems, calling Σ the complexity, or configurational entropy. An equivalent quantity is used in the study of random optimization problems by means of statistical physics techniques [22].

By following the analysis in [23], we shall call Eq. (1) the Einstein's likelihood principle; it is a fundamental relation in Einstein's theory of fluctuations [20]. Remarkably, in the context of superstatistics [10], Abe *et al.* [24] generalized Einstein's principle for nonequilibrium systems in the presence of (quenched) temperature fluctuations and using conditional entropies.

The aim of this paper is to provide a novel approach that relates, in a unique framework, classical information theory with both the notion of generalized entropy and Einstein's likelihood principle. Precisely, we shall show that an intrinsic *group-theoretical structure* is at the heart of the multiple connections among these foundational perspectives.

The physical root of this group structure relies in a recent generalization of the classical axiomatic formulation [25, 26], originally proposed by Shannon and by Khinchin to characterize the BG entropy. The first three postulates, nowadays called the Shannon–Khinchin (SK) axioms [1, 2], define some essential properties that any entropic functional S should satisfy. Let us consider the set \mathcal{P} of finite discrete probability distributions $P \in \mathcal{P}$, $P := \{p_1, \dots, p_W\}$, $W \in \mathbb{N} \setminus \{0\}$, $p_i \geq 0$, $\sum_{i=1}^W p_i = 1$. We can state the postulates as follows.

Continuity: The function $S(p_1, \dots, p_W)$ is continuous with respect to all its arguments.

Maximum principle: The function $S(p_1, \dots, p_W)$ is maximized by the uniform distribution.

Expansibility: Adding an event of zero probability does not change the value of the entropy, i.e., $S(p_1, \dots, p_W, 0) \equiv S(p_1, \dots, p_W)$.

Also, a fourth axiom, i.e. *additivity* with respect to the composition of two systems, was required. Under these assumptions, Khinchin proved that the only admissible entropy turns out to be the BG entropy $S^{\text{BG}}[P] := -\sum_{i=1}^W p_i \ln p_i$.

The additivity property was thought to be a sufficient condition for the *extensivity* of BG entropy, which in the formulation of Clausius is an essential requisite for thermodynamics. However, in the last decades it became evident that the two properties, i.e. additivity and extensivity, are unrelated [27]. Indeed, denoting by $W(N)$ the number of accessible states of a system with N particles, the BG entropy is not extensive, *on the uniform distribution*, if $W(N) \sim N^\alpha$ for a certain $\alpha \in \mathbb{R}^+$. This scaling is not atypical and it appears often in the framework of complex systems (see e.g. [28]). Recently, the non-extensivity of Boltzmann's entropy over a large class of probability distributions was proved. Surprisingly, Rényi's entropy can be extensive in the same contexts [29]. Additivity is therefore not an intrinsic physical requirement. At the same time, by renouncing the additivity postulate, new possibilities arise.

In the context of information theory, non-additive entropies provided useful tools, for example, in the study of quantum entanglement [9, 30]. However, the lack of additivity appears to be an important flaw if we want to interpret generalized entropies as *classical* measures of information [5, 14]. In the following, we shall assume that each system is in a given macrostate, corresponding to a set of possible configurations, or microstates; a

probability distribution function is defined over all allowed microstates. In this sense, each state is associated with a given probability distribution that, in our analysis, determines both the entropy and the information measure of the system. With a slight abuse of language, we will call the information measure obtained from this probability distribution as the information measure of the system itself. It is expected that, given two distributions associated with two independent systems (or with the results of two independent experiments), the corresponding total amount of information is nothing but the sum of the information measures obtained by each system (experiment). Moreover, any measurement or change of information in one of the two systems does not affect, nor is influenced by, any property of the other system, being the systems uncorrelated.

This property is certainly satisfied by Boltzmann's and Rényi's entropies; however, using non-additive entropic functionals *tout-court* it is not possible to fulfill the requirement above. Therefore, to preserve the interpretation of entropy as a generalized information measure $I(\mathcal{A})$ of a given system \mathcal{A} , we firstly postulate the general relation

$$I(\mathcal{A}) = f(S(\mathcal{A})), \quad (2)$$

where the information is assumed to be a function f of the entropy of the system only (hereafter we avoid to make explicit the dependence of I on the specific probability distribution P associated with a macrostate of \mathcal{A}).

Second, along the lines of [25, 26], we discuss an axiomatic formulation of entropy, generalizing the fourth SK axiom and requiring, instead of additivity, the *composability* property [31]. In this settings, we postulate that the entropy of a composite system be a function of the entropy of the two components only. This simple requirement is, however, rich with consequences. In particular, a specific algebraic structure appears, in which the composition operation plays the role of group operation. In the following, we will show that, for a large class of entropies, an additive information measure having the form of Eq. (2) is automatically inferred by the algebraic structure itself. The composition process of two or more systems is formalized as follows:

Composability axiom: Given two statistically independent systems \mathcal{A} and \mathcal{B} , each defined on a given probability distribution in \mathcal{P} , there exists a symmetric function $\Phi(x, y)$ such that

$$S(\mathcal{A} \cup \mathcal{B}) = \Phi(S(\mathcal{A}), S(\mathcal{B})). \quad (3)$$

Moreover, we require the associativity property

$$\Phi(\Phi(x, y), z) = \Phi(x, \Phi(y, z)) \quad (4)$$

and the relation $\Phi(x, 0) = x$.

In [32] axiomatic formulations of the notion of entropy, and in particular the composability property for non-independent systems have been discussed in a different framework.

We shall define *admissible entropies* as those satisfying the first three SK axioms and the *composability axiom*. The axiom contains fundamental requirements, for instance the existence of a zeroth law of thermodynamics. Also, when composing a system with another in a certainty state (zero entropy), the entropy of the composed system will remain unchanged. This natural generalization of the SK axiom allows an infinite number of admissible entropic forms [25, 26].

The classification and study of the properties of all functions $\Phi(x, y)$ satisfying the composability axiom has been performed in the context of formal group theory [33]. This branch of algebraic topology was established in the second half of the 20th century, starting from the work of Bochner [34]. In particular, there exists a universal group law, the *Lazard formal group*, essential to the discussion of composability.

More precisely, we shall consider a monotonic, strictly increasing function $G(t)$ admitting in \mathbb{R} a power series expansion of the form $G(t) = \sum_{k=0}^{\infty} a_k \frac{t^{k+1}}{k+1}$, where $\{a_k\}_{k \in \mathbb{N}}$ is a suitable real sequence, with $a_0 \neq 0$. This series, by means of the Lagrange inversion theorem, is invertible with respect to the composition (i.e., there exists a series G^{-1} such that $G^{-1}(G(t)) = t$). The function

$$\Phi(x, y) = G(G^{-1}(x) + G^{-1}(y)). \quad (5)$$

defines a group law. One can prove that, given a formal group law $\Phi(x, y)$, there exists a specific power series $G(t) = a_0 t + O(t^2)$ such that the law can be represented as in Eq. (5) [35].

Consequently, if a functional S satisfies the generalized SK axioms, then the associated composition law (3) can be realized in terms of the law (5) by means of a specific function G . The rich algebraic structure underlying the composability axiom naturally allows us to classify all entropies possessing a given composition law. Also, we can generate new examples, by varying the composition law adopted. For the trace-form family (i.e., the family of entropies having the structure $S[P] = \sum_i p_i f(p_i)$ for a sufficiently regular f), a natural general form is [26]

$$S_G[P] = \sum_{i=1}^W p_i G\left(\ln \frac{1}{p_i}\right). \quad (6)$$

For example, the celebrated Boltzmann entropy corresponds to the *additive group*:

$$\Phi(x, y) = x + y \Rightarrow G(t) = kt, \quad k \in \mathbb{R}^+, \quad (7)$$

the constant k being usually identified with the Boltzmann constant.

The *multiplicative group*

$$\Phi(x, y) = x + y + (1 - q)xy \Rightarrow G(t) = \frac{e^{a(1-q)t} - 1}{1 - q} \quad (8)$$

leads to a generalization of the Tsallis entropy S_q^T , defined for $q \in \mathbb{R}$ and $a \in \mathbb{R}^+$ as [26]

$$S_{a,q}[P] := \sum_{i=1}^W p_i \log_q \frac{1}{p_i^a}, \quad (9)$$

where $\log_q(x) := \frac{x^{1-q} - 1}{1 - q} \xrightarrow{q \rightarrow 1} \ln x$, $x \in \mathbb{R}^+$. Observe that

$$\lim_{a \rightarrow 1} S_{a,q}[P] = S_q^T[P] \quad \text{and} \quad \lim_{q \rightarrow 1} S_{a,q}[P] = a S^{\text{BG}}[P]. \quad (10)$$

Remarkably, the $S_{a,q}[P]$ entropy is the most general known trace-form entropy that is admissible according to the definition above, including the BG entropy and the Tsallis entropy as particular cases. All other entropies in the form (6) satisfy the composability property on the uniform distribution. Dropping out the trace-form hypothesis, new entropic forms are allowed [26]. For example, the Rényi entropy

$$S_\alpha^R[P] := \frac{\ln \sum_{i=1}^W p_i^\alpha}{1 - \alpha} \xrightarrow{\alpha \rightarrow 1} S^{\text{BG}}[P] \quad (11)$$

is an additive composable entropy.

Motivated by the previous discussion, we can propose now a notion of information measure that comes directly from the group-theoretical structure. To this regard, one should notice that the additivity property is lost for a very large class of composable entropies discussed above. However, we can overcome this difficulty by associating with each of them an information measure that is indeed additive on statistically independent systems. This measure is determined by the composition group itself, i.e., by the function G appearing in Eq. (5). Consequently, we propose the following main definition of a *group-theoretical generalized information measure*.

Given a composable entropy S_G , with a group law of the form (5), the information measure associated with S_G for any system \mathcal{A} corresponding to a given probability distribution is defined to be

$$I_G(\mathcal{A}) = G^{-1}(S_G(\mathcal{A})). \quad (12)$$

In the specific case of an entropy of trace-form class (6), we recover for our functional the expression of the *Kolmogorov-Nagumo mean* [36, 37]:

$$I_G(\mathcal{A}) = G^{-1}\left(\sum_{i=1}^W p_i G\left(\ln \frac{1}{p_i}\right)\right). \quad (13)$$

The information measure (12), however, is defined in the generic case of entropies that are composable and not only for trace-form entropies. We are ready now to present one of the main results of this paper.

Theorem. Let S_G be a composable entropy, with a group law defined by (5) for a certain function G . Then for two statistically independent systems \mathcal{A} and \mathcal{B} we have

$$I_G(\mathcal{A} \cup \mathcal{B}) = I_G(\mathcal{A}) + I_G(\mathcal{B}). \quad (14)$$

Moreover, I_G satisfies the following further properties.

Continuity: I_G is continuous with respect to its arguments;

Maximum principle: I_G is maximized on the uniform distribution;

Expansibility: The addition of a zero-probability event does not change the value of I_G .

Proof. Observe indeed that

$$\begin{aligned} I_G(\mathcal{A} \cup \mathcal{B}) &= \\ &= G^{-1}(S_G(\mathcal{A} \cup \mathcal{B})) = G^{-1}(\Phi(S_G(\mathcal{A}), S_G(\mathcal{B}))) \\ &= G^{-1}\{G[G^{-1}(S_G(\mathcal{A})) + G^{-1}(S_G(\mathcal{B}))]\} \\ &= I_G(\mathcal{A}) + I_G(\mathcal{B}). \end{aligned} \quad (15)$$

All the other properties of I_G derive from the properties of S_G imposed by the generalized SK axioms and from the strict monotonicity of G . \square

Therefore all composable entropies possess an associated information measure which is additive and can be constructed directly through the function G . Observe that the strict monotonicity of G (and therefore of G^{-1}) implies that $S_G(\mathcal{A}) < S_G(\mathcal{B}) \Rightarrow I_G(\mathcal{A}) < I_G(\mathcal{B})$, coherently with the fact that the entropic forms allowed by the axioms above do possess indeed an information content even if non-additive (see for example [9, 38] for the study of the Tsallis entropic form as information functional and for applications).

Observe also that the Rényi entropy fits naturally in our scheme. Indeed, for S_α^R , $G(t) = t$, then the associated information measure I_α^R coincides with the entropic functional, i.e.

$$S_\alpha^R(\mathcal{A}) \mapsto I_\alpha^R(\mathcal{A}) \equiv S_\alpha^R(\mathcal{A}). \quad (16)$$

The previous identity holds also in the $\alpha \rightarrow 1$ limit, i.e., in the BG case. In other words, the BG entropy and the Rényi entropy are stable with respect to definition (12): the associated group-theoretical information measure coincides with the corresponding entropy. If we consider instead the $S_{a,q}$ entropy, we see that

$$S_{a,q}(\mathcal{A}) \mapsto I_{a,q}(\mathcal{A}) \equiv S_{1+a(q-1)}^R(\mathcal{A}). \quad (17)$$

In particular, the information measure associated to S_q^T ($a \equiv 1$) is the Rényi entropic functional with parameter q . However, as explained above, our formalism goes

beyond the requirement of a Kolmogorov–Nagumo structure and it holds, indeed, for all nontrace-form admissible entropies, having a more general form for the corresponding information functional. For a large family of nontrace-form entropies (for example, for the entropic functionals discussed in [26]), the analytic expression of the information functional is of the type

$$S_G(\mathcal{A}) \mapsto I_G(\mathcal{A}) = \sum_i c_i S_{\alpha_i}^R(\mathcal{A}), \quad (18)$$

where the parameters c_i and α_i depend on the parameters appearing in the entropy S_G .

Finally, the presented group-theoretical approach allows us to generalize easily Einstein's principle, and to connect it with information theory in a natural way. Given an entropy S_G , whose associated group law is Eq. (5), we introduce the likelihood function

$$\mathcal{W}_G(\mathcal{A}) := e^{I_G(\mathcal{A})} \equiv e_G^{S_G(\mathcal{A})}, \quad (19)$$

where

$$e_G(x) := \exp[G^{-1}(x)]. \quad (20)$$

The motivation for this definition is twofold. First, it relates Einstein's likelihood function directly with a group-theoretical information measure. Second, it generalizes Einstein's relations [20, 23] in the case of independent systems for all composable entropies. Indeed, let S be a composable entropy, whose group law is given by Eq. (5). Then Einstein's likelihood principle (1) follows immediately from the additivity property of the information functional (12). In the case of the multiplicative group (8), we recover the likelihood function recently introduced in [23]. Once again, for generalized entropies that are composable only over the uniform distribution, just a weak formulation of the principle holds (needless to say, this situation is not very satisfactory).

In the light of the whole analysis of this paper, we conclude that the composability axiom allows a potentially fruitful interpretation of generalized entropies in information theory. Indeed, composable entropies both possess an information theoretical content and satisfy Einstein's principle, which is a crucial statement for any physical application of the notion of entropy. As a byproduct of our analysis, it emerges that also nontrace-form but composable entropies can play an important role in statistical mechanics.

Acknowledgments We thank Francesco Toppan and Constantino Tsallis for useful discussions. G.S. acknowledges the financial support of the John Templeton Foundation. The research of P.T. has been partly supported by the project FIS2015-63966, MINECO, Spain, and by the ICMAT Severo Ochoa project SEV-2015-0554 (MINECO).

* sicuro@cbpf.br

† p.tempesta@fis.ucm.es

- [1] C. Shannon, The Bell System Technical Journal **27**, 379 (1948).
- [2] A. Khinchin, *Mathematical Foundations of Information Theory*, Dover Books on Mathematics (Dover Publications, 1957).
- [3] E. T. Jaynes, Physical Review **106**, 620 (1957).
- [4] A. Rényi, in *Proc. 4th Berkeley Symp. Math. Statist. Probability* (1961) pp. 547–561.
- [5] A. Rényi, *Probability theory* (North-Holland, 1970).
- [6] J. Havrda and F. Charvát, Kybernetika **3**, 30 (1967).
- [7] C. Tsallis, Journal of statistical physics **52**, 479 (1988).
- [8] E. P. Borges and I. Roditi, Physics Letters A **246**, 399 (1998).
- [9] N. Canosa and R. Rossignoli, Physical Review Letters **88**, 170401 (2002).
- [10] C. Beck and E. Cohen, Physica A: Statistical mechanics and its applications **322**, 267 (2003).
- [11] R. Hanel, S. Thurner, and M. Gell-Mann, Proceedings of the National Academy of Sciences **108**, 6390 (2011).
- [12] P. Tempesta, Physical Review E **84**, 021121 (2011).
- [13] C. Tsallis and L. J. Cirto, The European Physical Journal C **73**, 1 (2013).
- [14] P. Jizba and T. Arimitsu, Annals of Physics **312**, 17 (2004).
- [15] P. Calabrese and J. Cardy, Journal of Physics A: Mathematical and Theoretical **42**, 504005 (2009).
- [16] A. Coser, E. Tonni, and P. Calabrese, Journal of Statistical Mechanics: Theory and Experiment **2014**, P12017 (2014).
- [17] P. Calabrese, F. H. Essler, M. Andreas, *et al.*, Journal of Statistical Mechanics: Theory and Experiment **2014**, P09025 (2014).
- [18] P. Calabrese, P. Le Doussal, and S. N. Majumdar, Physical Review A **91**, 012303 (2015).
- [19] E. Cohen, Physica A: Statistical Mechanics and its Applications **305**, 19 (2002).
- [20] A. Einstein, *Annalen der Physik* **338**, 1275 (1910).
- [21] M. Mézard and G. Parisi, Journal of Statistical Physics **111**, 1 (2003).
- [22] V. Bapst, L. Foini, F. Krzakala, G. Semerjian, and F. Zamponi, Physics Reports **523**, 127 (2013).
- [23] C. Tsallis and H. J. Haubold, EPL (Europhysics Letters) **110**, 30005 (2015).
- [24] S. Abe, C. Beck, and E. G. Cohen, Physical Review E **76**, 031102 (2007).
- [25] P. Tempesta, Annals of Physics **365**, 180 (2016).
- [26] P. Tempesta, arXiv:1507.07436 (2015).
- [27] H. Touchette, Physica A: Statistical Mechanics and its Applications **305**, 84 (2002).
- [28] R. Hanel and S. Thurner, EPL (Europhysics Letters) **96**, 50003 (2011).
- [29] H. Bergeron, E. Curado, J. Gazeau, and L. M. Rodrigues, Physica A **441**, 23 (2016).
- [30] M. Berta, M. Christandl, R. Colbeck, J. M. Renes, and R. Renner, Nature Physics **6**, 659 (2010).
- [31] C. Tsallis, *Introduction to Nonextensive Statistical Mechanics: Approaching a Complex World* (Springer New York, 2009).
- [32] C. Beck, Contemporary Physics **50**, 495 (2009).
- [33] M. Hazewinkel, *Formal groups and applications*, Vol. 78 (Elsevier, 1978).
- [34] S. Bochner, Annals of Mathematics, 192 (1946).
- [35] This set of properties holds for group laws on zero-characteristic fields. For further details, see [26].
- [36] A. Kolmogorov, in *Atti della R. Accademia Nazionale dei Lincei. Memorie della Classe di scienze fisiche, matematiche e naturali*, Vol. 12 (1930) pp. 388–391.
- [37] M. Nagumo, in *Japanese journal of mathematics: transactions and abstracts*, Vol. 7 (The Mathematical Society of Japan, 1930) pp. 71–79.
- [38] C. Tsallis, Physical Review E **58**, 1442 (1998).